

Fullname: Phạm Văn Phúc

Student's ID: 522H0068

Homework 2. Mining frequent patterns and Association Rules

	i1	i2	i3	i4	i5	i6	i7	i8
T1	1	0	0	0	0	0	1	1
T2	1	1	0	0	0	1	1	1
T3	1	1	0	0	0	1	1	0
T4	1	0	0	0	0	0	1	1
T5	0	0	1	1	1	1	0	1
T6	1	0	0	1	1	0	0	0

a. Find frequent patterns at **minsupp=0,3**

Frequent patterns are itemsets with support ≥ 2 . I will use the Apriori method to find frequent itemsets, starting with 1-itemsets, then expanding to larger sizes.

Step 1: Check single items first (support ≥ 2)

Count the number of occurrences of each item:

- {i1}: 5 (T1, T2, T3, T4, T6) \rightarrow supp = $5 \geq 2 \rightarrow$ Frequent
- {i2}: 2 (T2, T3) \rightarrow supp = $2 \geq 2 \rightarrow$ Frequent
- {i3}: 1 (T5) \rightarrow supp = $1 < 2 \rightarrow$ Not frequent
- {i4}: 2 (T5, T6) \rightarrow supp = $2 \geq 2 \rightarrow$ Frequent
- {i5}: 2 (T5, T6) \rightarrow supp = $2 \geq 2 \rightarrow$ Frequent
- {i6}: 3 (T2, T3, T5) \rightarrow supp = $3 \geq 2 \rightarrow$ Frequent
- {i7}: 4 (T1, T2, T3, T4) \rightarrow supp = $4 \geq 2 \rightarrow$ Frequent
- {i8}: 5 (T1, T2, T4, T5, T6) \rightarrow supp = $5 \geq 2 \rightarrow$ Frequent

So frequent 1-itemsets: {i1}, {i2}, {i4}, {i5}, {i6}, {i7}, {i8}

Step 2: Check frequent 2-itemsets (support ≥ 2)

Combine pairs of frequent 1-itemsets and check for support:

- {i1, i2}: (T2, T3) = 2 \rightarrow Frequent
- {i1, i4}: (T6) = 1 \rightarrow Not frequent
- {i1, i5}: (T6) = 1 \rightarrow Not frequent

- $\{i1, i6\}: (T2, T3) = 2 \rightarrow \text{Frequent}$
- $\{i1, i7\}: (T1, T2, T3, T4) = 4 \rightarrow \text{Frequent}$
- $\{i1, i8\}: (T1, T2, T4, T6) = 4 \rightarrow \text{Frequent}$
- $\{i2, i4\}: 0 \rightarrow \text{Not frequent}$
- $\{i2, i5\}: 0 \rightarrow \text{Not frequent}$
- $\{i2, i6\}: (T2, T3) = 2 \rightarrow \text{Frequent}$
- $\{i2, i7\}: (T2, T3) = 2 \rightarrow \text{Frequent}$
- $\{i2, i8\}: (T2) = 1 \rightarrow \text{Not frequent}$
- $\{i4, i5\}: (T5, T6) = 2 \rightarrow \text{Frequent}$
- $\{i4, i6\}: (T5) = 1 \rightarrow \text{Not frequent}$
- $\{i4, i8\}: (T5, T6) = 2 \rightarrow \text{Frequent}$
- $\{i5, i6\}: (T5) = 1 \rightarrow \text{Not frequent}$
- $\{i5, i8\}: (T5, T6) = 2 \rightarrow \text{Frequent}$
- $\{i6, i7\}: (T2, T3) = 2 \rightarrow \text{Frequent}$
- $\{i6, i8\}: (T2, T5) = 2 \rightarrow \text{Frequent}$
- $\{i7, i8\}: (T1, T2, T4, T5) = 4 \rightarrow \text{Frequent}$

So frequent 2-itemsets: $\{i1, i2\}, \{i1, i6\}, \{i1, i7\}, \{i1, i8\}, \{i2, i6\}, \{i2, i7\}, \{i4, i5\}, \{i4, i8\}, \{i5, i8\}, \{i6, i7\}, \{i6, i8\}, \{i7, i8\}$

Step 3: Check frequent 3-itemsets (support ≥ 2)

Combine pairs of frequent 2-itemsets that share at least 1 item and check:

- $\{i1, i2, i6\}: (T2, T3) = 2 \rightarrow \text{Frequent (from } \{i1, i2\} \text{ và } \{i1, i6\})$
- $\{i1, i2, i7\}: (T2, T3) = 2 \rightarrow \text{Frequent (from } \{i1, i2\} \text{ và } \{i2, i7\})$
- $\{i1, i6, i7\}: (T2, T3) = 2 \rightarrow \text{Frequent (from } \{i1, i6\} \text{ và } \{i6, i7\})$
- $\{i1, i7, i8\}: (T1, T2, T4, T5) = 4 \rightarrow \text{Frequent (from } \{i1, i7\} \text{ và } \{i7, i8\})$
- $\{i2, i6, i7\}: (T2, T3) = 2 \rightarrow \text{Frequent (from } \{i2, i6\} \text{ và } \{i6, i7\})$
- $\{i4, i5, i8\}: (T5, T6) = 2 \rightarrow \text{Frequent (from } \{i4, i5\} \text{ và } \{i5, i8\})$
- $\{i6, i7, i8\}: (T2, T5) = 2 \rightarrow \text{Frequent (from } \{i6, i7\} \text{ và } \{i6, i8\})$

(Other combinations like $\{i1, i2, i8\}$ only have $\text{supp} = 1$, which is not enough)

Frequent 3-itemsets: $\{i1, i2, i6\}, \{i1, i2, i7\}, \{i1, i6, i7\}, \{i1, i7, i8\}, \{i2, i6, i7\}, \{i4, i5, i8\}, \{i6, i7, i8\}$

Step 4: Find Frequent 4-Itemsets

Combine from frequent 3-itemsets:

- $\{i1, i2, i6, i7\}$: $(T2, T3) = 2 \rightarrow$ Frequent (from $\{i1, i2, i6\}$ và $\{i2, i6, i7\}$)
- $\{i1, i6, i7, i8\}$: $(T2) = 1 \rightarrow$ Not frequent
- $\{i1, i2, i7, i8\}$: $(T2) = 1 \rightarrow$ Not frequent

Frequent 4-itemsets: $\{i1, i2, i6, i7\}$

Step 5: Check 5-Itemsets

- $\{i1, i2, i6, i7, i8\}$: $(T2) = 1 \rightarrow$ Not frequent

There is no set larger than 4-itemsets that satisfies.

Conclusion: All Frequent Patterns

- 1-itemsets: $\{i1\}, \{i2\}, \{i4\}, \{i5\}, \{i6\}, \{i7\}, \{i8\}$
- 2-itemsets: $\{i1, i2\}, \{i1, i6\}, \{i1, i7\}, \{i1, i8\}, \{i2, i6\}, \{i2, i7\}, \{i4, i5\}, \{i4, i8\}, \{i5, i8\}, \{i6, i7\}, \{i6, i8\}, \{i7, i8\}$
- 3-itemsets: $\{i1, i2, i6\}, \{i1, i2, i7\}, \{i1, i6, i7\}, \{i1, i7, i8\}, \{i2, i6, i7\}, \{i4, i5, i8\}, \{i6, i7, i8\}$
- 4-itemsets: $\{i1, i2, i6, i7\}$

b. Find max-patterns at **minsupp=0,3**

Max-patterns are frequent itemsets that are not subsets of any other frequent itemset. We check from the largest set down:

- **4-itemsets:**
 - $\{i1, i2, i6, i7\}$: There is no superset larger than \rightarrow Maximal
- **3-itemsets** (check if is a subset of $\{i1, i2, i6, i7\}$):
 - $\{i1, i2, i6\}$: Subset of $\{i1, i2, i6, i7\} \rightarrow$ Not maximal
 - $\{i1, i2, i7\}$: Subset of $\{i1, i2, i6, i7\} \rightarrow$ Not maximal
 - $\{i1, i6, i7\}$: Subset of $\{i1, i2, i6, i7\} \rightarrow$ Not maximal
 - $\{i1, i7, i8\}$: Not a subset \rightarrow Maximal
 - $\{i2, i6, i7\}$: Subset of $\{i1, i2, i6, i7\} \rightarrow$ Not maximal
 - $\{i4, i5, i8\}$: Not a subset \rightarrow Maximal
 - $\{i6, i7, i8\}$: Not a subset \rightarrow Maximal

The 2-itemsets and 1-itemsets are both subsets of larger sets, so no further testing is necessary.

Conclusion: Max-Patterns

- {i1, i2, i6, i7}
 - {i1, i7, i8}
 - {i4, i5, i8}
 - {i6, i7, i8}
- c. Find all of the association rules (**minsupp=0,3** and **minconf=1**) that came from max-patterns in question b.

Generate rules $A \rightarrow B$ from max-patterns, with:

- $\text{supp}(A \cup B) \geq 2$
- $\text{conf}(A \rightarrow B) = \frac{\text{supp}(A \cup B)}{\text{supp}(A)} = 1$

From Max-Pattern {i1, i2, i6, i7} (supp = 2, T2, T3)

- Support for subsets:
 - {i1}: 5, {i2}: 2, {i6}: 3, {i7}: 4
 - {i1, i2}: 2, {i1, i6}: 2, {i1, i7}: 4, {i2, i6}: 2, {i2, i7}: 2, {i6, i7}: 2
 - {i1, i2, i6}: 2, {i1, i2, i7}: 2, {i1, i6, i7}: 2, {i2, i6, i7}: 2
- Law:
 - {i1, i6, i7} \rightarrow {i2}: $2/2 = 1$
 - {i2, i6, i7} \rightarrow {i1}: $2/2 = 1$
 - {i1, i2} \rightarrow {i6, i7}: $2/2 = 1$
 - {i2, i6} \rightarrow {i1, i7}: $2/2 = 1$
 - {i2, i7} \rightarrow {i1, i6}: $2/2 = 1$
 - {i6, i7} \rightarrow {i1, i2}: $2/2 = 1$
 - (Other laws such as {i1} \rightarrow {i2, i6, i7} have $\text{conf} = 2/5 < 1$, not satisfied)

From Max-Pattern {i1, i7, i8} (supp = 4, T1, T2, T4, T5)

- Support for subsets:
 - {i1}: 5, {i7}: 4, {i8}: 5
 - {i1, i7}: 4, {i1, i8}: 4, {i7, i8}: 4
- Law:
 - {i1, i7} \rightarrow {i8}: $4/4 = 1$
 - {i1, i8} \rightarrow {i7}: $4/4 = 1$
 - {i7, i8} \rightarrow {i1}: $4/4 = 1$

From Max-Pattern {i4, i5, i8} (supp = 2, T5, T6)

- Support for subsets:
 - $\{i4\}: 2, \{i5\}: 2, \{i8\}: 5$
 - $\{i4, i5\}: 2, \{i4, i8\}: 2, \{i5, i8\}: 2$
- Law:
 - $\{i4, i5\} \rightarrow \{i8\}: 2/2 = 1$
 - $\{i4, i8\} \rightarrow \{i5\}: 2/2 = 1$
 - $\{i5, i8\} \rightarrow \{i4\}: 2/2 = 1$
 - $(\{i8\} \rightarrow \{i4, i5\})$ have $\text{conf} = 2/5 < 1$, not satisfied)

From Max-Pattern $\{i6, i7, i8\}$ (supp = 2, T2, T5)

- Support for subsets:
 - $\{i6\}: 3, \{i7\}: 4, \{i8\}: 5$
 - $\{i6, i7\}: 2, \{i6, i8\}: 2, \{i7, i8\}: 4$
- Law:
 - $\{i6, i7\} \rightarrow \{i8\}: 2/2 = 1$
 - $\{i6, i8\} \rightarrow \{i7\}: 2/2 = 1$
 - $(\{i7, i8\} \rightarrow \{i6\})$ have $\text{conf} = 2/4 < 1$, not satisfied)

Conclusion:

- $\{i1, i6, i7\} \rightarrow \{i2\}$
- $\{i2, i6, i7\} \rightarrow \{i1\}$
- $\{i1, i2\} \rightarrow \{i6, i7\}$
- $\{i2, i6\} \rightarrow \{i1, i7\}$
- $\{i2, i7\} \rightarrow \{i1, i6\}$
- $\{i6, i7\} \rightarrow \{i1, i2\}$
- $\{i1, i7\} \rightarrow \{i8\}$
- $\{i1, i8\} \rightarrow \{i7\}$
- $\{i7, i8\} \rightarrow \{i1\}$
- $\{i4, i5\} \rightarrow \{i8\}$
- $\{i4, i8\} \rightarrow \{i5\}$
- $\{i5, i8\} \rightarrow \{i4\}$
- $\{i6, i7\} \rightarrow \{i8\}$
- $\{i6, i8\} \rightarrow \{i7\}$